2016-08-03 Metadata Curation Meeting notes

Date

📅 03 Aug 2016

Attendees

- Kathleen Carr
- Dana Shum
- Tyler Stevens
- Alicia Aleman
- Nathan James
- Kaylin Bugbee
- Jeanne' le Roux

Goals

- Review goals and status of Curation team
- Determine steps to get to a 'unified curation process'

Discussion Items

Item	Who	Notes
Set context for this meeting by reviewing meeting minutes from previous meetings of the group	all	- The original workflow drawn by Katie after the May Metadata Curation Summit may need some revision (see Next Meeting Discussion) - On that original workflow, we have only been working on the ARC and GCMD parts of the flow, not on the CMR or HDF parts
Current status of ARC curation of GHRC records	Kathleen Carr Kaylin Bugbee Jeanne' le Roux	- ARC is tasked to curate all (~350) of the GHRC records, regardless of BEDI status - ARC is curating each collection record, plus one representative granule record for each collection - ARC has currently completed 225 Collection records and 225 Granule records - ARC is capturing curation results in a spreadsheet which is passed on to the GHRC DAAC for action (not using JIRA yet as a problem reporting tool) - GHRC DAAC is tracking remediation progress in their own copy of the ARC spreadsheet
Current status of GCMD curation of GHRC records	Kathleen Carr Tyler Stevens	- GCMD is tasked to curate the selected 25 collection records from GHRC - GCMD has completed curation of those 25 records
What are the goals of our team?	all	- ARC and GCMD individually curate 25 collection records from GHRC (NOTE: these records are native ECHO 10) - Use that experience to compare and contrast curation processes and come up with a 'unified curation process' - Go forward using the 'unified curation process' to curate CMR metadata records at the remaining DAACs
What are the steps to creating a 'unified curation process'?	all	- Gather existing documentation on ARC and GCMD curation processes - Analyze existing processes and determine overlaps, unique checks, best practices (GCMD has already started to document this; use GCMD spreadsheet as a starting point) - Determine single best curation process for records and fields - Determine unified output format and repository for curation results to give to DAACs for remedial action (e.g. spreadsheet, JIRA tickets)

Where do we document and maintain the 'unified curation process'?	all	- Need a common repository - github may not work because GCMD code is not Open Source - Dana will check with ESDIS regarding setting up an Earthdata Code Collaborative (ECC) repository
Once we have arrived at a 'unified curation process', should we have a Pathfinder DAAC to try out that process?	all	- GHRC, if we are still working through those records; DAAC #2 if not; - DAAC # 2 candidates might be LPDAAC or JPL/PODAAC, since these DAACs have a large number of BEDI records
Which records should be curated at the remaining DAACs?	all	Not clear from our July meeting whether we made a final decision to curate the BEDI datasets first; need to confirm with ESDIS

Action items

- ☑ Jeanne' le Roux Choose a single GHRC record (concept Id) from the 25 that both teams curated, to be used as an example for comparing GCMD and ARC processes and results
- ☑ Dana Shum check with ESDIS re which repository to use for 'unified script ' (probably ECC; since github has OpenSource issues)
- ☐ Kathleen Carr Draft revised curation process workflow for ESDIS review
- ☑ Kathleen Carr Schedule our next meeting to begin creating a unified curation process, using the single GHRC record selected by Jeanne as our working example
- ☐ Kathleen Carr Produce list of BEDI datasets at GHRC for reference